

Citation for published version:

Margulis, AV, Palmsten, K, Andrade, SE, Charlton, RA, Hardy, JR, Cooper, WO & Hernández-Díaz, S 2015, 'Beginning and duration of pregnancy in automated health care databases: Review of estimation methods and validation results', *Pharmacoepidemiology and Drug Safety*, vol. 24, no. 4, pp. 335-342.
<https://doi.org/10.1002/pds.3743>

DOI:

[10.1002/pds.3743](https://doi.org/10.1002/pds.3743)

Publication date:

2015

Document Version

Peer reviewed version

[Link to publication](#)

This is the peer reviewed version of the following article: Margulis, A. V., Palmsten, K., Andrade, S. E., Charlton, R. A., Hardy, J. R., Cooper, W. O., & Hernández-Díaz, S. (2015). Beginning and duration of pregnancy in automated health care databases: Review of estimation methods and validation results. *Pharmacoepidemiology and Drug Safety*, 24(4), 335-342, which has been published in final form at <http://dx.doi.org/10.1002/pds.3743>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results

Running head: Pregnancy beginning and duration in automated data

Authors

Andrea V Margulis, MD, ScD ¹

Kristin Palmsten, ScD ^{2, 3}

Susan E Andrade, ScD ⁴

Rachel A Charlton, PhD ⁵

Janet R Hardy, PhD ⁶

William O Cooper, MD, MPH ⁷

Sonia Hernández-Díaz, MD, DPH ³

¹ RTI Health Solutions, Barcelona, Spain

² Department of Pediatrics, University of California, San Diego, La Jolla, California, USA

³ Department of Epidemiology, Harvard School of Public Health, Boston, USA

⁴ Meyers Primary Care Institute, University of Massachusetts Medical School, Worcester, MA, USA

⁵ Department of Pharmacy and Pharmacology, University of Bath, Bath, UK

⁶ Department of Epidemiology and Biostatistics, University of South Florida, and ECC Population Health Group, St Petersburg, USA

⁷ Departments of Pediatrics and Preventive Medicine, Vanderbilt University School of Medicine,
Nashville, Tennessee, USA

Corresponding author

Andrea V Margulis

andreamargulis@post.harvard.edu

Travessera de Gracia, 56, atico 1

08006 Barcelona

Spain

Tel.: +34 693 822 166

Fax: +34 934 142 610

Keywords

- Administrative data
- Beginning of pregnancy
- Claims data
- Duration of pregnancy
- Electronic medical records
- Last menstrual period

Key points

- Several methods to overcome the lack of information in health care databases on the date of beginning of pregnancy, or its duration, have been used in epidemiologic research.
- These methods vary in complexity, data element requirements, limitations, and accuracy.

Funding sources

Kristin Palmsten was supported by Training Grant T32HD060454 in Reproductive, Perinatal and Pediatric Epidemiology from the National Institute of Child Health and Human Development, National Institutes of Health. Rachel Charlton is currently funded by a European Union Seventh Framework Programme grant (HEALTH-F5-2011-260598). Rachel Charlton has also received funding from GlaxoSmithKline, Norgine and Novartis Vaccines and Diagnostics. Janet Hardy has consulted for Teva Pharmaceutical Industries, GlaxoSmithKline, and AbbVie. Sonia Hernández-Díaz was supported by the Agency for Healthcare Research and Quality (AHRQ) (Grant R01HS018533). Dr. Hernández-Díaz has consulted for GSK-Biologics and Novartis. The Pharmacoepidemiology Program at the Harvard School of Public Health receives funding from Pfizer, Millennium, Phrma, Bayer and Asisa.

Conflicts of interest: none

Word count: 3,220

Prior presentations

An earlier version of this work was presented in the symposium "Electronic Healthcare Databases for Pregnancy Research: Panacea or Trojan Horse?" at the 28th International Conference on Pharmacoepidemiology and Therapeutic Risk Management held in Barcelona

(August 25, 2012). We would like to acknowledge the contribution from the symposium attendees, whose useful comments were incorporated in this review.

Ethics

This is a review of published papers and does not require ethical approval.

Acknowledgements

We would like to thank one of the anonymous reviewers, who generously provided very helpful suggestions throughout this manuscript's peer review process.

Abbreviations

FDA US Food and Drug Administration

ICD International Classification of Diseases

GPRD General Practice Research Database (currently Clinical Practice Research Datalink, CPRD)

NPV negative predictive value

PPV positive predictive value

THIN The Health Improvement Network

This is the accepted version of the following article: Margulis AV, Palmsten K, Andrade SE, Charlton RA, Hardy JR, Cooper WO, *et al.* Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results. *Pharmacoepidemiology and Drug Safety*. 2015;24(4):335-342. which has been published in final form at <http://onlinelibrary.wiley.com/doi/10.1002/pds.3743/full>

Authors' accepted version

Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results

ABSTRACT

Purpose

To describe methods reported in the literature to estimate the beginning or duration of pregnancy in automated health care data, and to present results of validation exercises where available.

Methods

Papers reporting methods for determining the beginning or duration of pregnancy were identified based on Pubmed searches, by consulting investigators with expertise in the field and by reviewing conference abstracts and reference lists of relevant papers. From each paper or abstract, we extracted information to characterize the study population, data sources, and estimation algorithm. We then grouped these studies into categories reflecting their general methodological approach.

Results

Methods were classified into 5 categories: (1) methods that assign a uniform duration for all pregnancies, (2) methods that assign pregnancy duration based on preterm-delivery or health care related codes, or codes for other pregnancy outcomes, (3) methods based on the timing of prenatal care, (4) methods based on birth weight, and (5) methods that combine elements from 2 and 3. Validation studies evaluating these methods used varied approaches, with results

generally reporting on the mistiming of the start of pregnancy, incorrect estimation of the duration of pregnancy or misclassification of drug exposure during pregnancy or early pregnancy.

Conclusions

In the absence of accurate information on the beginning or duration of pregnancy, several methods of varying complexity are available to estimate them. Validation studies have been performed for many of them and can serve as a guide for method selection for a particular study.

Beginning and duration of pregnancy in automated health care databases: review of estimation methods and validation results

INTRODUCTION

Research on medications and other exposures in pregnancy is critical for public health. As exposures and pregnancy outcomes in need of evaluation are often uncommon, automated databases may be the only data sources able to provide reasonably precise estimates of exposure prevalence or relative risks in a cost efficient and timely manner. This is because they collect information on the health care provided to large populations, and are readily available.

These data sources also present challenges,¹⁻⁴ particularly regarding the accurate ascertainment of the beginning of pregnancy, which is needed to assess the gestational age at the time of exposure or other events. Further, there are specific periods when disruptions of the physiological processes in the development of the embryo, fetus or related structures may increase the risk of an adverse outcome. Identifying these periods requires the beginning of pregnancy be known. In addition, outcomes with definitions based on gestational age, such as preterm delivery or fetal growth, must also be evaluated with reference to the beginning of pregnancy.

Information on the onset of pregnancy is usually absent from claims data and not always present in electronic medical record databases.⁵ However, both types of automated data generally have information on delivery or birth dates, or proxies for them (e.g., delivery hospital

admission date).^{6, 7} Some data sources can be linked to birth certificates (e.g., Kaiser Permanente, Tennessee Medicaid), delivery hospitalization records (e.g., British Columbia, Ontario, Quebec) or medical records to obtain information on the date of conception, date of last menstrual period, expected due date, or gestational age at birth. But there are situations in which the linkage is not possible, the information on the beginning or duration of pregnancy is not available in the linked data, or the public health need for results is pressing and data linkage processing times are too long. Then, methods to overcome the lack of information on the start or length of pregnancy must be implemented.

In this paper, we review methods that have been used to determine the beginning of pregnancy or the first trimester, or to estimate the duration of pregnancy, as well as validation studies that have evaluated their accuracy.

LITERATURE SEARCH AND GROUPING OF METHODS

The description of methods used for estimating pregnancy onset or duration is generally not reflected in papers' titles, abstracts or key words. For that reason, we did not conduct a systematic literature search to identify studies for inclusion in the present review. Instead, we conducted multiple Pubmed searches using varied strategies, consulted investigators with expertise in research on drug safety in pregnancy with automated databases, examined conference abstract books and reviewed the reference section in papers already identified as relevant for this review.

We grouped methods reflecting their general approach in: (1) methods that assign a uniform duration of pregnancy, (2) methods that assign pregnancy duration based on preterm-delivery or related codes, or codes for other pregnancy outcomes, (3) methods based on the timing of prenatal care, (4) methods based on birth weight, and (5) methods that combine elements from 2 and 3.

From each manuscript or abstract, we abstracted information to characterize the study population, data sources, and estimation algorithm, highlighting findings from the earliest paper we found that used each methodology or the one that provided the most detailed description. We describe the methods' ease of implementation and note which pregnancies would be excluded and which would have their duration incorrectly estimated. Then we discuss which exposures and outcomes the methods would be incapable of addressing and other limitations. Common themes are discussed at the end. Results from validation studies are presented in the online supplementary material. The online supplementary material also contains expanded information on each method.

DESCRIPTION OF METHODS, STRENGTHS AND LIMITATIONS

Methods, and their strengths and limitations, are summarized in Table 1 and the text below.

While some methods estimated a start date, others estimated the duration of pregnancy (the beginning of pregnancy can be back-calculated from the pregnancy end date). Validation studies evaluating each method also used varied approaches. Results generally report on the

mistiming of the start of pregnancy, misclassification of drug exposure in pregnancy or early pregnancy, or incorrect estimation of the duration of pregnancy.

1. Methods that assign a uniform duration of pregnancy

These methods only require information on the pregnancy end date and are the easiest to implement: the beginning of pregnancy is estimated by subtracting the estimated duration of pregnancy from the pregnancy end date (e.g., date of delivery or proxies such as date of delivery hospital admission) in all pregnancies (Table 1 and Supplementary Table 1). The duration of pregnancy has generally been assumed to be between 270 and 280 days.

The method's main limitation is that the estimated pregnancy duration is several weeks too long in all preterm⁸ and early term deliveries and a few days to weeks too short in all late term and post-term deliveries. Since many outcomes of interest are associated with a shorter pregnancy (e.g., some congenital malformations, preeclampsia), the duration of pregnancy might be incorrectly estimated among those pregnancies with the outcome. Consequently, gestational exposures among cases might be differentially misclassified. As a consequence, relative risk estimates for these outcomes can be biased toward or against the null. For example, for exposures that decrease after conception, cases would have an artificially high apparent prevalence of exposure in early pregnancy (Figure 1, top panel).

In response to this, pregnancies with an increased risk of preterm delivery (e.g., pregnancies with hypertension or cervical incompetence) can be analyzed separately⁹ or excluded.^{10, 11} This has important consequences: first, data from both mother and offspring are often needed to identify pregnancies at high risk of preterm delivery. Second, once these pregnancies have been excluded, the study population decreases in size and may no longer be representative of the overall population. Third, the distribution of exposures directly or indirectly associated (whether harmful or protective) with conditions selected for the identification of likely preterm pregnancies (e.g., drugs to treat hypertension, diabetes, or procedures to treat cervical incompetence) would be distorted in the study population. Moreover, preterm delivery and outcomes associated with preterm delivery cannot be studied using these methods.

2. Methods that assign pregnancy duration based on preterm-delivery or related codes, or codes for other pregnancy outcomes

Commonly used health care coding systems (e.g., International Classification of Diseases [ICD] 9, ICD-10, Read codes) include codes for preterm delivery, preterm-related maternal or infant conditions, or use of related health care services. Codes may provide varying levels of detail, such as the relatively vague Read code Q11..11, “Baby born premature” or the more specific Read code 635B.00, “Baby premature 36 weeks”. Relevant information may reside in maternal or offspring, outpatient or inpatient data. The methods we present in this section use pregnancy-specific data to make pregnancy-specific estimations of pregnancy duration and are straight forward to apply (Table 1 and Supplementary Table 2). One method¹² assumed all

pregnancies with codes for miscarriage lasted 180 days, those with codes for preterm delivery lasted 210 days, etc. Another method¹³ timed the conception window 252 to 287 days prior to delivery for singleton births, and 238 to 273 for births of multiple infants, and assumed a 4-month long first trimester spanning from the earlier bound of the conception window to 91 days after the later bound. Another method¹⁴ used information on the date of the last menstrual period and gestational age at birth in codes and free text in electronic medical records when available. The remaining pregnancies were assumed to reach 280 days unless there was information on preterm delivery. Another study⁸ created an indicator of preterm delivery using codes for short gestation or preterm labor and assigned pregnancies with the indicator a duration of 245 days and those without it, 273 days (these were the preferred durations based on validation results). Another method¹⁵ used multiple codes for preterm-related disorders, birth weight and post-natal care to create an indicator of delivery at 237 days (less than 34 completed weeks) or earlier. Another approach¹⁶ assumed a duration of pregnancy equal to the upper limit of the range of gestational age at birth when the information was available in codes (e.g., ICD-9 code 765.28, “35-36 completed weeks of gestation”), 245 days for preterm-related codes with no gestational age specification, or 270 days for the remaining pregnancies. An algorithm that incorporates ICD-9 codes for prolonged and post-term pregnancies has been proposed but, to our knowledge, has not yet been used.¹⁷

The ability to identify preterm deliveries avoids their exclusion from the study population and allows estimation of the duration of preterm pregnancies with less error than the methods previously described. Exposures and outcomes related to preterm delivery can thus be studied.

Still, with some of the methods, the earliest deliveries would be assigned too long a pregnancy duration. Post-term deliveries have most often been grouped with term deliveries. This is less of a concern, because post-term deliveries are generally not much longer than term pregnancies, as labor is typically induced at 41 or 42 completed weeks.¹⁸

3. Methods based on timing of prenatal health care

Three approaches made use of information from prenatal health services utilization present in maternal records to make pregnancy-specific estimations (Table 1 and Supplementary Table 3). One study¹⁹ used the date of the first booking, ambulatory or inpatient prenatal-care related code to define two time windows, one before and one after the code date. The first time window likely included the date of conception and the earliest weeks of pregnancy. The second window estimated the first trimester of pregnancy. Women with late initiation of prenatal care would have their two windows mistimed. Therefore, pregnancies with less than 7 months of prenatal care were excluded. Another study⁸ assumed ambulatory prenatal screening tests that are recommended for a certain period of pregnancy (e.g., measurement of alpha fetoprotein, gestational weeks 11-14) are performed in the middle of the interval, and used the date of the test to estimate the date of the beginning of pregnancy. Information from multiple prenatal screening tests per pregnancy were combined through linear regression to obtain a single estimate of pregnancy duration for each pregnancy. Another study²⁰ used a similar approach but instead of combining the information from multiple tests per pregnancy, hierarchically selected a single prenatal test and based the duration estimate on it.

These approaches rely on the assumption that pregnant women receive a recommended schedule of prenatal health care and will perform better in settings where prenatal care is available and highly standardized. Women with late initiation of prenatal care or who are screened outside the typical schedule for other reasons, however, would have their windows or their beginning of pregnancy mistimed. The first of these three methods¹⁹ identified and excluded women who had a short prenatal follow-up. This exclusion reduced the study population and may specifically have depleted its shorter gestation higher risk pregnancies. If not excluded, pregnancies with late entry into the health care system would have had their conception date and first trimester estimated as too late. The second method⁸ dealt with non-standard prenatal care by only utilizing the first occurrence of each screening test in each pregnancy and making use of all patterns of prenatal care found in the study population, at the cost of a substantial complexity in the analysis.

Limitations of the first method¹⁹ include conception falling within some unknown distribution of time before the first pregnancy marker and the application of exclusion criteria that omit a sizeable proportion of the population. Its main strength is that it is very straight forward to apply. The main limitations of the second method⁸ are the relatively intensive data management required, and the assumptions on the similarity of prenatal care between the study population in which the algorithm was developed and the population on which it would be applied. Its most important strength is its improved performance among preterm deliveries over other available methods. The third method²⁰ is easier to implement, but would similarly

depend on assumptions of comparable health care patterns. These three methods may be problematic in the study of exposures or outcomes related to non-standard care. For example, women who conceive using assisted reproductive technology may begin their health care very early and have earlier pregnancy markers than is typical. To accommodate this, the time windows used in the first method ¹⁹ could be redefined for earlier pregnancy markers. A remaining limitation is that diverse populations may access prenatal care differently, including groups at high risk for adverse outcomes.

4. Method based on birth weight

This method, ²¹ applied to pregnancies ending in a live or stillbirth with gestational age information at birth missing from birth certificates, is easy to implement (Table 1 and Supplementary Table 4). Using the race- and year-specific distribution of weight for gestational age at birth (derived from study pregnancies with complete data), this method assumes the birth weight is located on the median of the distribution and imputes the missing gestational age at birth with the value that corresponds to the recorded birth weight.

A limitation of this method is that exposures or conditions that may affect growth (e.g., smoking, gestational diabetes) or outcomes related to growth or weight cannot be studied. For example, term infants small for their gestational age may be misclassified as preterm, possibly leading to differential misclassification of exposure (Figure 1, middle panel). Birth weight needs to be available, as well as summary statistics on birth weight for births of similar gestational

age, race and year of birth with known date of last menstrual period. Conceivably, the latter could be replaced with external data on a comparable population, if available.

5. Methods that combine elements from #2 and #3

One method²² implemented a hierarchical algorithm that used information from hospital discharge records, laboratory tests, emergency department admissions and other data in a stepwise fashion to derive the pregnancy start date for each pregnancy (Table 1 and Supplementary Table 5). When none of this information was available, the algorithm resorted to national vital statistic figures for the median gestational age by outcome for seven outcome categories: very preterm live birth, preterm live birth, full-term live birth, spontaneous abortion, therapeutic abortion, stillbirth or ectopic pregnancy. In another study,²³ authors adapted the algorithm so that it would not require information from pregnancy laboratory tests. The adapted algorithm includes information from the health plan diagnosis, procedure, and birth files.

An additional study²⁴ used a range of pregnancy related codes to determine the pregnancy duration in electronic medical records (algorithm details were presented in the poster associated with the referenced abstract). Codes were categorized and used in the following order of priority: a record for the estimated date of delivery, a record of the last menstrual period, gestational age time-related records (e.g. “A/N 20 week examination”, “Baby premature 36 weeks”) and records relating to a preterm or post-term delivery which did not specify the

gestational age, which were set to a default duration of 36 weeks and 41 weeks respectively.

For all remaining pregnancies a default pregnancy duration of 280 days for a delivery, including stillbirths, and 70 days for a pregnancy loss was assigned. A simplified version of this algorithm was used in a later publication .²⁵

The main advantage of these algorithms is that they make use of multiple sources of information. Imputing gestational age from national vital statistics ensures there will be no pregnancies whose duration is not estimable, as long as the pregnancy outcome is known and vital statistics for that outcome are available. Disadvantages include those discussed in the sections on methods that assign pregnancy duration based on preterm-delivery or related codes, and on methods based on timing of prenatal health care.

DISCUSSION

A number of methods to estimate the beginning or duration of pregnancy with satisfactory validation results have been proposed and used. It is reasonable to use all available and reliable information and minimize blanket imputations such as a fixed duration of pregnancy. The shorter the expected duration of exposure (e.g., an antibiotic therapy episode), the more important it is to obtain a precise (and correct) estimate of the pregnancy start date or duration. Chronic exposures are less vulnerable to misclassification^{8, 16, 25, 26} than episodic exposures (mismeasurement of cumulative exposures may occur however, as shown in Figure 1). Likewise, outcomes with a short risk period (e.g., congenital malformations) or that change quickly over pregnancy (e.g., gestational weight gain or fetal growth) warrant precise estimates.

Prenatal and postnatal care evolve and new screening tests and treatments become available. Estimation methods will need to be adapted to reflect current practices, which may differ across country, regions, facilities or health plans, and may vary over time in studies with a long study period. With current efforts to delay delivery until term or near term, methods that assume term pregnancy may be more precise for recent cohorts than for historical ones. Many of the methods described here require linkage of maternal and infants' health care records, or maternal health care records and infants' birth certificates to select the study population (e.g., to exclude preterm infants) or to extract other information needed for estimations. The linkage process may be complex and result in the loss of non-linkable potential study subjects, decreasing the study size. Further, if excluded subjects are not a random sample of the population, the study population may not be representative of the source population.²⁷ On the other hand, preterm delivery codes may be less accurate in studies that only use claims from maternal data than in studies that also use infant's data.²⁸

The list of papers, methods and validation results presented here (validation results are provided in the online supplementary material) is comprehensive but not necessarily exhaustive. We report detail to provide context to methods and validation results, but more information can be found in the original publications. The direction and amount of bias associated with the use of different estimation methods will vary depending on the study question and data characteristics. As mentioned, validation studies took very different

approaches, which limits the comparability of the results but provides a wide frame for the appreciation of methods' caveats and strengths.

Spontaneous or elective terminations, miscarriages and stillbirths

Methods and validation results presented in this review are based on live births, with some exceptions.^{12, 21, 22, 24} Some methods could be generalized to pregnancies that were terminated or ended in a spontaneous abortion or a stillbirth. For example, a method based on the first record of prenatal care¹⁹ could be used in such cases, as long as it is reasonable to assume that the initiation of prenatal care in those pregnancies has a comparable distribution to that described for term and close-to-term pregnancies.

Conclusions

In the absence of accurate information on the beginning or duration of pregnancy, several methods are available to estimate them. Methods vary in sophistication, information needed to implement them, and applications for which they are useful. While broad assignments of a fixed gestational age should be avoided, the most complex algorithms may not be optimal for a given study either. Validation results can serve as a guide in the selection of a method for a particular study.

REFERENCES

1. Hernandez-Diaz S. Prescription of medications during pregnancy: accidents, compromises, and uncertainties. *Pharmacoepidemiol Drug Saf.* 2006 Sep; **15**(9): 613-7. DOI: 10.1002/pds.1308
2. Andrews EB, Tennis P. Promise and pitfalls of administrative data in evaluating pregnancy outcomes. *Pharmacoepidemiol Drug Saf.* 2007 Nov; **16**(11): 1181-3. DOI: 10.1002/pds.1499
3. Grimes DA. Epidemiologic research using administrative databases: garbage in, garbage out. *Obstet Gynecol.* 2010 Nov; **116**(5): 1018-9. DOI: 10.1097/AOG.0b013e3181f98300
4. Palmsten K, Hernandez-Diaz S. Can nonrandomized studies on the safety of antidepressants during pregnancy convincingly beat confounding, chance, and prior beliefs? *Epidemiology.* 2012 Sep; **23**(5): 686-8. DOI: 10.1097/EDE.0b013e318258fbb2
5. Manson JM, McFarland B, Weiss S. Use of an automated database to evaluate markers for early detection of pregnancy. *American journal of epidemiology.* 2001 Jul 15; **154**(2): 180-7.
6. Raebel MA, Ellis JL, Andrade SE. Evaluation of gestational age and admission date assumptions used to determine prenatal drug exposure from administrative data. *Pharmacoepidemiol Drug Saf.* 2005 Dec; **14**(12): 829-36. DOI: 10.1002/pds.1100
7. Garbe E, Suling M, Kloss S, *et al.* Linkage of mother-baby pairs in the German Pharmacoepidemiological Research Database. *Pharmacoepidemiol Drug Saf.* 2011 Mar; **20**(3): 258-64. DOI: 10.1002/pds.2038
8. Margulis AV, Setoguchi S, Mittleman MA, *et al.* Algorithms to estimate the beginning of pregnancy in administrative databases. *Pharmacoepidemiol Drug Saf.* 2013 Jan; **22**(1): 16-24. DOI: 10.1002/pds.3284

9. Andrade SE, Gurwitz JH, Davis RL, *et al.* Prescription drug use in pregnancy. *Am J Obstet Gynecol.* 2004 Aug; **191**(2): 398-407. DOI: 10.1016/j.ajog.2004.04.025
10. Grisso JA, Strom BL, Cosmatos I, *et al.* Diagnostic ultrasound in pregnancy and low birthweight. *Am J Perinatol.* 1994 Jul; **11**(4): 297-301. DOI: 10.1055/s-2007-994597
11. Andrade SE, Raebel MA, Morse AN, *et al.* Use of prescription medications with a potential for fetal harm among pregnant women. *Pharmacoepidemiol Drug Saf.* 2006 Aug; **15**(8): 546-54. DOI: 10.1002/pds.1235
12. Alonso A, Jick SS, Olek MJ, *et al.* Recent use of oral contraceptives and the risk of multiple sclerosis. *Arch Neurol.* 2005 Sep; **62**(9): 1362-5. DOI: 10.1001/archneur.62.9.1362
13. Cole JA, Modell JG, Haight BR, *et al.* Bupropion in pregnancy and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf.* 2007 May; **16**(5): 474-84. DOI: 10.1002/pds.1296
14. Petersen I, Gilbert RE, Evans SJ, *et al.* Pregnancy as a major determinant for discontinuation of antidepressants: an analysis of data from The Health Improvement Network. *J Clin Psychiatry.* 2011 Jul; **72**(7): 979-85. DOI: 10.4088/JCP.10m06090blu
15. Eworuke E, Hampp C, Saidi A, *et al.* An algorithm to identify preterm infants in administrative claims data. *Pharmacoepidemiol Drug Saf.* 2012 Jun; **21**(6): 640-50. DOI: 10.1002/pds.3264
16. Li Q, Andrade SE, Cooper WO, *et al.* Validation of an algorithm to estimate gestational age in electronic health plan databases. *Pharmacoepidemiol Drug Saf.* 2013 May; **22**(5): 524-32. DOI: 10.1002/pds.3407
17. Kawai AT, Li L, Andrade SE, *et al.* MINI-SENTINEL CBER/PRISM SURVEILLANCE PROTOCOL, INFLUENZA VACCINES AND BIRTH OUTCOMES, Final Draft Posted for Public Comment 2013 March 10, 2013. Available from: http://www.mini-sentinel.org/work_products/PRISM/Mini-Sentinel_PRISM_Influenza-Vaccines-and-Birth-Outcomes-Protocol.pdf.

18. Gulmezoglu AM, Crowther CA, Middleton P. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev*. 2006; (4): CD004945. DOI: 10.1002/14651858.CD004945.pub2
19. Hardy JR, Leaderer BP, Holford TR, *et al*. Safety of medications prescribed before and during early pregnancy in a cohort of 81,975 mothers from the UK General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2006 Aug; **15**(8): 555-64. DOI: 10.1002/pds.1269
20. Kharbanda EO, Vazquez-Benitez G, Shi WX, *et al*. Assessing the safety of influenza immunization during pregnancy: the Vaccine Safety Datalink. *Am J Obstet Gynecol*. 2012 Sep; **207**(3 Suppl): S47-51. DOI: 10.1016/j.ajog.2012.06.073
21. Piper JM, Mitchel EF, Jr., Ray WA. Presumptive eligibility for pregnant Medicaid enrollees: its effects on prenatal care and perinatal outcome. *American journal of public health*. 1994 Oct; **84**(10): 1626-30.
22. Hornbrook MC, Whitlock EP, Berg CJ, *et al*. Development of an algorithm to identify pregnancy episodes in an integrated health care delivery system. *Health services research*. 2007 Apr; **42**(2): 908-27. DOI: 10.1111/j.1475-6773.2006.00635.x
23. Naleway AL, Gold R, Kurosky S, *et al*. Identifying pregnancy episodes, outcomes, and mother-infant pairs in the Vaccine Safety Datalink. *Vaccine*. 2013 Jun 12; **31**(27): 2898-903. DOI: 10.1016/j.vaccine.2013.03.069
24. Snowball J, De Vries C. Determination of Pregnancy on the General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2007; **16**: S118.
25. Charlton RA, Weil JG, Cunningham MC, *et al*. Comparing the General Practice Research Database and the UK Epilepsy and Pregnancy Register as tools for postmarketing teratogen surveillance: anticonvulsants and the risk of major congenital malformations. *Drug Saf*. 2011 Feb 1; **34**(2): 157-71. DOI: 10.2165/11584970-000000000-00000

26. Toh S, Mitchell AA, Werler MM, *et al.* Sensitivity and specificity of computerized algorithms to classify gestational periods in the absence of information on date of conception. *American journal of epidemiology*. 2008 Mar 15; **167**(6): 633-40. DOI: 10.1093/aje/kwm367
27. Palmsten K, Huybrechts KF, Mogun H, *et al.* Harnessing the Medicaid Analytic eXtract (MAX) to Evaluate Medications in Pregnancy: Design Considerations. *PloS one*. 2013; **8**(6): e67405. DOI: 10.1371/journal.pone.0067405
28. Andrade SE, Scott PE, Davis RL, *et al.* Validity of health plan and birth certificate data for pregnancy research. *Pharmacoepidemiol Drug Saf*. 2013 Jan; **22**(1): 7-15. DOI: 10.1002/pds.3319
29. Jick SS, Terris BZ. Anticonvulsants and congenital malformations. *Pharmacotherapy*. 1997 May-Jun; **17**(3): 561-4.
30. Grisso JA, Carson JL, Feldman HI, *et al.* Epidemiological pitfalls using Medicaid data in reproductive health research. *The Journal of maternal-fetal medicine*. 1997 Jul-Aug; **6**(4): 230-6. DOI: 10.1002/(SICI)1520-6661(199707/08)6:4<230::AID-MFM9>3.0.CO;2-K
31. Andrade SE, McPhillips H, Loren D, *et al.* Antidepressant medication use and risk of persistent pulmonary hypertension of the newborn. *Pharmacoepidemiol Drug Saf*. 2009 Mar; **18**(3): 246-52. DOI: 10.1002/pds.1710
32. Charlton RA, Cunningham MC, de Vries CS, *et al.* Data resources for investigating drug exposure during pregnancy and associated outcomes: the General Practice Research Database (GPRD) as an alternative to pregnancy registries. *Drug Saf*. 2008; **31**(1): 39-51. DOI: 3114 [pii]
33. Drinkard CR, Shatin D, Clouse J. Postmarketing surveillance of medications and pregnancy outcomes: clarithromycin and birth malformations. *Pharmacoepidemiol Drug Saf*. 2000 Dec; **9**(7): 549-56. DOI: 10.1002/pds.538

34. Palmsten K, Setoguchi S, Margulis AV, *et al.* Elevated risk of preeclampsia in pregnant women with depression: depression or antidepressants? *American journal of epidemiology*. 2012 May 15; **175**(10): 988-97. DOI: 10.1093/aje/kwr394
35. Devine ST, Hammad TA. Last Menstrual Period Estimation Techniques Compared to Physician Recorded Last Menstrual Period Dates in the General Practice Research Database [abstract]. *Pharmacoepidemiol Drug Saf*. 2008; **17**: S298.
36. Cole JA, Ephross SA, Cosmatos IS, *et al.* Paroxetine in the first trimester and the prevalence of congenital malformations. *Pharmacoepidemiol Drug Saf*. 2007 Oct; **16**(10): 1075-85. DOI: 10.1002/pds.1463
37. Palmsten K, Hernandez-Diaz S, Kuriya B, *et al.* Use of disease-modifying antirheumatic drugs during pregnancy and risk of preeclampsia. *Arthritis care & research*. 2012 Nov; **64**(11): 1730-8. DOI: 10.1002/acr.21807
38. Margulis AV, Abou-Ali A, Strazzeri MM, *et al.* Use of selective serotonin reuptake inhibitors in pregnancy and cardiac malformations: a propensity-score matched cohort in CPRD. *Pharmacoepidemiol Drug Saf*. 2013 Sep; **22**(9): 942-51. DOI: 10.1002/pds.3462
39. Cooper WO, Willy ME, Pont SJ, *et al.* Increasing use of antidepressants in pregnancy. *Am J Obstet Gynecol*. 2007 Jun; **196**(6): 544 e1-5. DOI: 10.1016/j.ajog.2007.01.033
40. Piper JM, Baum C, Kennedy DL. Prescription drug use before and during pregnancy in a Medicaid population. *Am J Obstet Gynecol*. 1987 Jul; **157**(1): 148-56.
41. Ray WA, Gigante J, Mitchel EF, Jr., *et al.* Perinatal outcomes following implementation of TennCare. *JAMA*. 1998 Jan 28; **279**(4): 314-6.
42. Cooper WO, Hernandez-Diaz S, Arbogast PG, *et al.* Major congenital malformations after first-trimester exposure to ACE inhibitors. *N Engl J Med*. 2006 Jun 8; **354**(23): 2443-51. DOI: 10.1056/NEJMoa055202

43. Ban L, Tata LJ, West J, *et al.* Live and non-live pregnancy outcomes among women with depression and anxiety: a population-based study. *PloS one*. 2012; **7**(8): e43462. DOI: 10.1371/journal.pone.0043462
44. Charlton RA, Hutchison A, Davis KJ, *et al.* Asthma management in pregnancy. *PloS one*. 2013; **8**(4): e60247. DOI: 10.1371/journal.pone.0060247
45. Sammon CJ, Snowball J, McGrogan A, *et al.* Evaluating the hazard of foetal death following H1N1 influenza vaccination; a population based cohort study in the UK GPRD. *PloS one*. 2012 Dec 20; **7**(12): e51734. DOI: 10.1371/journal.pone.0051734
46. Hardy JR, Holford TR, Hall GC, *et al.* Strategies for identifying pregnancies in the automated medical records of the General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2004 Nov; **13**(11): 749-59. DOI: 10.1002/pds.935
47. Piper JM, Ray WA, Griffin MR, *et al.* Methodological issues in evaluating expanded Medicaid coverage for pregnant women. *American journal of epidemiology*. 1990 Sep; **132**(3): 561-71.
48. Devine S, West S, Andrews E, *et al.* The identification of pregnancies within the general practice research database. *Pharmacoepidemiol Drug Saf*. 2010 Jan; **19**(1): 45-50. DOI: 10.1002/pds.1862

Table 1. Summary of Methods

Method	Strengths	Limitations	Validated?
1. Methods that assign a uniform duration of pregnancy ^{9-11, 29-34}	Easy to implement with limited data on mother and infant	Estimated duration of pregnancy too long in preterm deliveries If preterm deliveries are excluded, distortion of distribution of preterm-related exposures and events may occur Estimated duration of pregnancy too short in post-term deliveries	Yes ^{6, 8, 26, 35}
2. Methods that assign pregnancy duration based on preterm-delivery or related codes, or codes for other pregnancy outcomes ^{8, 12-16, 34, 36-38}	Exclusion of preterm births is avoided. Studies of exposures and outcomes related to preterm birth are feasible.	Estimated duration of pregnancy too long in the earliest preterm deliveries Estimated duration of pregnancy too short in post-term deliveries	Yes ^{8, 15, 16}
3. Methods based on timing of prenatal health care ^{8, 19, 20}	Estimates are based on prospective prenatal care received by each pregnant woman	Estimated duration of pregnancy possibly incorrect in pregnancies that receive non-standard health care	Yes ^{8, 20, 26, 35}
4. Method based on birth weight ^{21, 39-42}	Estimates are based on information from birth certificates, considered reliable	Infants who were born too large or too little for their gestational age and race would have their duration of gestation misestimated Summary information on birth weight by gestational age and race is required	No
5. Methods that combine elements of #2 and #3 ^{20, 22-25, 43-45}	See #2 and #3	See #2 and #3	Yes ^{22, 23}

Validation results and greater detail on methods are presented in the online supplementary material.

The first part of this online supplementary material presents selected validation results for each group of methods. The second part consists of 5 tables with expanded details on the methods and validation studies.

PART I. Validation results

1. Methods That Assign a Uniform Duration of Pregnancy

In a validation study that assumed a duration of pregnancy of 270 days in all gestations, the authors compared results with the gold standard, (self-reported duration of pregnancy) and stratified analyses into preterm/term pregnancies. The sensitivity for exposure (antiinfective use in the first trimester) was 65.8% in preterm pregnancies and 92.5% in term gestations.²⁶ In another study, with an estimated gestational age at birth of 273 days,⁸ the estimated beginning of pregnancy was within 2 weeks of the gold standard in 99.1% of the term pregnancies. However, 64% of preterm pregnancies had an estimate that was 2-4 weeks too early. In 4%, the estimate was more than 4 weeks too early. In relation to misclassification of exposure, compared to estimates based on the admission date as a proxy for delivery date and a pregnancy duration of 270 days, first-trimester drug use overall was 1.3% higher when the duration of pregnancy was estimated from the delivery date and the gestational age at birth information obtained from a birth registry.⁶

2. Methods That Assign Pregnancy Duration Based on Preterm-delivery or Related Codes, or Codes for Other Pregnancy Outcomes

Two validation studies compared the pregnancy duration estimated with methods in this group to the pregnancy duration in discharge records for the delivery hospitalization or birth certificate files, stratified by preterm status. In one of them, the estimated and the recorded

duration of pregnancy were within 2 weeks of each other in 75% of preterm and 99% of term deliveries.⁸ When looking at first-trimester use of medications, a chronic exposure (fluoxetine) based on algorithm-derived gestational age at birth had both sensitivity and positive predictive value (PPV) of 97%, and both specificity and negative predictive value (NPV) of 100%, relative to exposure timing determined from birth-certificate derived gestational age at birth. In the case of a shorter exposure (first-trimester use of amoxicillin), the sensitivity was 93%, PPV 92%, specificity and NPV 100%.¹⁶

3. Methods Based on Timing of Prenatal Health Care

A validation study showed that, using the first of these methods, 77% of the initial study population met eligibility criteria related to duration of prenatal follow-up. In this group, the sensitivity of exposure (first-trimester use of antiinfectives) was 59%, and the specificity, 98%.

²⁶ In a validation study, the second method estimated the gestational age at birth within 2 weeks of the gold standard from hospital discharge records in 75% of preterm deliveries, and in 93% of term deliveries. Although this represents a marginal improvement for preterm deliveries over simpler methods based on the presence of preterm codes, this method has performed worse than simpler methods among term deliveries.⁸ The third method has been reported to result in a mean difference of less than 1 day compared to birth-certificate gestational age and to be over 97% accurate in assigning first-trimester exposure to influenza vaccine.²⁰

4. Method Based on Birth Weight

No validation studies found.

5. Methods That Combine Elements From #2 and #3

In one study,²² 81% of pregnancies matching on outcome date agreed on gestational age at birth within 2 weeks, and 13% agreed within 15-28 days. In the second study,²³ algorithm-derived gestational ages were within 30 days of the corresponding gestational ages

documented in medical charts for 98% of live births, 83% of spontaneous abortions, 85% of therapeutic abortions, and 82% of other pregnancy outcomes.

Authors' accepted version

PART II. Details of methods

TABLE 1. Methods That Assign a Uniform Duration of Pregnancy

Publication	Study population	Data source	Estimation method	Comments	Publications which used related methods
Grisso, 1994 ¹⁰	Singleton pregnancies (maternal age 15; 55 years old) with one or more inpatient delivery codes, one or more health-related visits in the 9 months or more before delivery and a linked child in the database were eligible. One pregnancy per woman was retained. Each case of low-birth weight was matched with 4 control pregnancies on maternal age at delivery and year of delivery. For subgroup analyses, pregnancies with codes for pregnancy complications within 270 days prior to delivery or codes for preterm delivery were excluded.	1982-1987 Health care claims from Medicaid from a single state, USA	The duration of pregnancy was assumed to be 40 weeks. Gestational weeks and trimesters were back calculated from delivery.	40 weeks was the window selected for exposure assessment. For the purpose of subgroup analyses, codes for pregnancy complications were searched for in the 270 days prior to delivery.	Grisso, 1997 ³⁰
Jick, 1997 ²⁹	Women who delivered a liveborn and had entered the database at least 1 year before delivery.	1988-1993 Electronic medical records (GPRD), UK	"The date of conception was calculated from gestational age and date of birth. If the gestational age was unknown, it was assumed to be 280 days".	No description of the mother-child matching process.	Drinkard, 2000 ³³ Charlton, 2008 ³² Andrade, 2009 ³¹ Palmsten, 2012 ³⁴

Andrade, 2004 ⁹	Women who delivered an infant in a hospital with continuous enrollment and prescription drug coverage for at least 1 year prior to delivery. Excluded those with no evidence of prenatal care within 270 days of delivery (live births and stillbirths).	1996-2000 Health care claims from 8 health managed organizations in 7 states, USA	Pregnancy duration was assumed to be 270 days, with 3 90-day trimesters of pregnancy. The presence of diagnostic codes for conditions associated with preterm delivery (present in 15% of the deliveries) was used to stratify analyses.	In a perinatal database maintained by one of the plans, the mean gestational length was 273 days and the median was 275 days.
Andrade, 2006 ¹¹	Women who delivered an infant in a hospital with continuous enrollment and prescription drug coverage for at least 1 year before delivery. Excluded those without evidence of prenatal care in the 270 days prior to the delivery date. Retained only the first eligible delivery per woman. Pregnancies with diagnostic codes for conditions associated with a preterm delivery, or prescriptions for ovulation stimulants in the 270 preceding delivery were excluded.	1996-2000 Health care claims from 8 health managed organizations in 7 states, USA	Pregnancies were assumed to have started 270 days prior to delivery date.	For women who were included in the study and in a perinatal database maintained by one of the plans, the mean gestational length was 273 days and the median was 275 days.

Validation studies				
Publication	Study population	Data source	Methods	Results

Raebel, 2005 ⁶	<p>Claims-based dataset: Female members > 15 years old admitted to a contracting hospital for delivery, with continuous enrollment and prescription drug coverage in pregnancy.</p> <p>Registry-based dataset: Female members > 15 years old with prescription drug coverage in pregnancy identified in a related birth registry that records deliveries at ≥24 weeks of gestation in the health care system.</p>	<p>1997-2000</p> <p>Health care claims from a single health care plan in a single state, USA</p>	<p>In the claims-based dataset, pregnancies were assumed to have a duration of 270 days. The first trimester was defined as days 270 to 181 before admission for delivery (proxy for the delivery date). In the registry-based dataset, the first trimester was defined as the initial 90 days of gestation calculated from the delivery date and registry-recorded gestational age at birth. The two datasets were compared. Proportions of drug use in the first trimester and in the last 60 days of pregnancy among women present in both datasets were calculated.</p>	<p>Of 541 deliveries identified in one dataset only, in 45% the reason was that the 270 day or the admission as a proxy for delivery assumption did not hold, often combined with enrollment- or coverage-related eligibility criteria. The difference in proportions of drug use (any drug, FDA pregnancy category D, or category X) at any time in pregnancy, in the first trimester of pregnancy, or in the last 60 days of pregnancy comparing both datasets ranged between 0.02% and 1.04%. The difference in proportions relative to the drug use observed in the claims-based dataset ranged between 1.28% and 13.67%.</p>
Toh, 2008 ²⁶	<p>3177 infants without malformations and their mothers.</p>	<p>1998-2006</p> <p>Slone Epidemiology Center Birth Defects Study, data from a single state</p>	<p>This study replicated Andrade's method ¹¹ as closely as possible, excluding conditions associated with preterm delivery. The first trimester of pregnancy was defined as the first 90 days of pregnancy starting from 270 days prior to delivery. Exposure to antiinfectives in the first trimester was estimated and compared to the exposure to the same drugs using Slone Epidemiology Center's information on the date of last menstrual period as the gold standard. Sensitivity and specificity of the estimated first-trimester exposure to antiinfectives were reported.</p>	<p>From 3177 women in the data source, 1949 were eligible following Andrade's study eligibility criteria. The sensitivity (95% CI) for first-trimester exposure to antiinfectives was 92.1% (87.6%; 95.3%); the specificity was 99.5% (99.1%; 99.8%). When, instead of the original eligibility criteria, the 3177 women were included and stratified by term/preterm birth, the sensitivity for terms was 92.5% (89.3%; 95.0%); for preterms, it was 65.8% (48.7%; 80.4%). Specificity remained above 97% in all analyses. ^a</p>

Devine, 2008 ^{35 b}	417,946 live births	1987-2006 Electronic medical records (GPRD), UK	The date of last menstrual period was back calculated from 301,384 expected delivery dates. Those with high certainty ("sure" in the data field for expected delivery date certainty) or those with a date of last menstrual period close to the one in the additional clinical data were retained. Then, the date of last menstrual period was estimated as 280 days prior to the delivery date in all deliveries.	52,073 physician-based dates of last menstrual period were identified and linked to 30,240 estimated dates. The absolute mean difference (95% CI) between the physician-based and the estimated date of last menstrual period was 6.4 days (6.3; 6.5). For deliveries with delivery date within 7 days of the expected delivery date, it was 3.7 days (3.7; 3.8); and, for the rest, 10.7 (10.8; 10.8). ^c
Margulis, 2013 ⁸	Hospital live births in province's perinatal database with 365+280 days of outpatient coverage prior to the delivery date and gestational age at birth between 20 and 44 completed weeks.	1998-2007 Health care claims from single payer province-run health care system, Canada	The duration of pregnancy was estimated as 280, or 273 days, and compared to the clinical gestational age at birth (reference) in the delivery hospitalization discharge record. Results are reported stratified by gestational age at birth <37 (preterm deliveries) or ≥37 completed weeks (term deliveries) in hospital records.	The estimated gestational age at birth was within 2 weeks of the clinical gestational age at birth in 0% of preterm deliveries for an estimated duration of pregnancy of 280 or 273 days; in term deliveries, the figures were 93.3% for an estimated duration of pregnancy of 280 days, and 99.1% for an estimated duration of pregnancy of 273 days.

CI: confidence interval; FDA: US Food and Drug Administration; GPRD: General Practice Research Database (currently Clinical Practice Research Datalink, CPRD)

Note that we provide examples of publications which used the methods presented rather than a full list

^a Results from sensitivity analyses with varying duration of exposure and on relative risks for hypothetical outcomes are also presented.

^b Abstract

^c Sic

TABLE 2. Methods That Assign Pregnancy Duration Based on Preterm-delivery or Related Codes, or Codes for Other Pregnancy Outcomes

Publication	Study population	Data source	Estimation method	Comments	Publications which used related methods
Alonso, 2005 ¹²	Women aged < 50 years old with a multiple sclerosis diagnosis with ≥ 3 years of continuous enrollment before the first symptoms of multiple sclerosis, and up to 10 matched controls with the same eligibility criteria except for the multiple sclerosis diagnosis. Matching was based on age, practice and date of joining the practice.	1993-2000 Electronic medical records (GPRD), UK	The duration of pregnancy was assumed to be 270 days for term deliveries and stillbirths, 210 days for preterm deliveries, 180 days for miscarriages, and 120 days for induced abortions.	The authors report that other realistic imputations of the duration of pregnancy yielded similar results (not shown).	
Cole, 2007 ¹³	Women 12 - 49 years old dispensed antidepressants who had a live birth, with 1 year of continuous enrollment before their delivery date.	1995-2004 Health care claims from several health plans from several regions, US	Conception window: Singleton births: 287 days to 252 days before delivery. Birth of multiple infants: 273 days to 238 days before delivery. Estimated first trimester: from the earliest probable conception date to 91 days after the last probable conception date. No mention of how multiple versus singleton pregnancies were identified.	A sensitivity analysis in data from women whose date of last menstrual period was available was conducted; the original odds ratios were corrected for misclassification of the first trimester; estimates did not change (not shown).	Cole, 2007 ³⁶
Petersen, 2011 ¹⁴	Pregnancies in women who had a live birth and were registered with a general practice for at least 6 months before the beginning of pregnancy and were still registered at the time of delivery	1992 -2006 Electronic medical records (THIN), UK	“The duration of pregnancy was determined from information on the gestational age for the baby at birth, the Read code entries for the last menstrual period, and associated free text. If this information was not available and if there was no information in the notes of a preterm birth, the length of the pregnancy was imputed as 280 days. This method was used for 31% of the pregnancies.”		

Eworuke, 2012 ¹⁵	Infants with continuous eligibility for fee-for-service benefits from birth to the earliest of 3 months of age or death ("live births delivered between 15 and 50 weeks"), linked to birth or death certificates	1999-2004 Medicaid claims from 2 states, USA	Evaluated the sensitivity and PPV of 5-digit ICD-9 codes within code 765 to identify gestational age at birth up to 34 completed weeks. Also, used logistic regression with duration of pregnancy data from birth certificates as outcome, and 53 parameters related to birth weight, neonatal disorders and post-natal care from claims within 3 months of birth to estimate a prematurity score for each infant. Prematurity was defined as gestational age at birth < 34 completed weeks. Selected the score with PPV≥90% that minimized false positives and assessed accuracy of the score. Up to 5-digit ICD-9 codes required. Assumed all pregnancies with a score above the threshold had duration < 34 weeks.	
Margulis, 2013 ⁸	Hospital live births in province's perinatal database with 365+280 days of outpatient coverage prior to the delivery date and gestational age at birth between 20 and 44 completed weeks.	1998-2007 Health care claims from single payer province-run health care system, Canada	Preterm deliveries were identified via a preterm status indicator (presence of claims with 3 or 4-digit ICD-9 or 10 codes for disorders related to short gestation or early labor) in the first 60 days after delivery. Infants with codes for preterm were assigned a duration of gestation of 245 days and those without were assigned 273 days.	Palmsten, 2012 ³⁴ Palmsten, 2012 ³⁷
Li, 2013 ¹⁶	Pregnancies that ended in a live birth in women aged 15-45 years, with gestational age in the linked birth certificate between 20 and 45 weeks and compatible with the infant's birth weight. For drug exposure assessments, it was further required that women had continuous plan enrollment with pharmacy benefits from 100 days prior to pregnancy through delivery.	2001-2007 Health care claims from several health plans in several regions, USA	Gestational age at birth was assigned as: the upper limit of code description in pregnancies with codes specifying gestational age at birth; 245 days in pregnancies with codes for preterm delivery with no specification of gestational age at birth; 270 days for the remaining pregnancies.	Birth certificates had a last-menstrual-period derived gestational age, and a clinical or obstetric estimate; the authors found the difference was small and used the last menstrual period estimate in their analysis. The validity of birth certificate data for some of this data source had been previously validated.

Validation studies				
Publication	Study population	Data source	Methods	Results
Eworuke, 2012 ¹⁵	See above	See above	See above. The sensitivity of ICD-9 codes 765.21-765.27 individually and overall, relative to delivery at gestational age up to 34 weeks, and the sensitivity, specificity and PPV for the prematurity score were estimated in each state.	The sensitivity of individual codes 765.21-765.27 ranged between 11.4% and 15.5% in Texas, and from 19.9% and 29.7% in Florida. The sensitivity of the codes overall was 12.5% in Texas and 25.7% in Florida. The sensitivity of the prematurity score was 46.8% in Texas and 52.6% in Florida; its specificity, 99.9% and 99.8%; and its PPV, 82.2% and 91.7%, respectively.
Li, 2012 ¹⁶	See above	See above	See above. Summary statistics of the estimated gestational age are contrasted to those of the gold standard gestational age at birth from birth certificates. Drug use was assessed from dispensed prescriptions, adding 14 days to the days of supply. The sensitivity, specificity, PPV and NPV of drug use are reported. ^a	Based on the algorithm, the prevalence of preterm delivery was 8.4%, while based on birth certificate data it was 15.3%. Among preterm deliveries (based on birth certificate data), 77.1% had an estimated duration of pregnancy within 2 weeks of the clinical gestational age at birth. Among term deliveries, the figure was 77%. First trimester exposure to fluoxetine based on the algorithm-derived gestational at birth had sensitivity and PPV of 97%, and specificity and NPV of 100%. For amoxicillin, the sensitivity was 93%, PPV 92%, specificity and NPV 100%.
Margulis, 2013 ⁸	See above	See above	See above. Sensitivity, specificity, PPV and NPV of the preterm status indicator are presented. Agreement of estimated duration of pregnancy and clinical gestational age at birth are reported stratified by gestational age at birth <37 (preterm deliveries) or ≥37 completed weeks (term deliveries) in hospital records.	The preterm status indicator had a prevalence of 8.5%, compared to 6.9% from vital statistics for the province. The sensitivity was 91% (95% CI 91%; 91%); the specificity, 98% (98% ; 98%); the PPV, 74%; and the NPV, 99%. Among preterm deliveries (defined based on hospital-discharge data), 74.8% had an estimated duration of pregnancy within 2 weeks of the clinical gestational age at birth. Among term

deliveries, the figure was 99.1%. Other assumptions were evaluated and showed lower percentages of agreement.

CI: confidence interval; ICD: International Classification of Diseases; GPRD: General Practice Research Database (currently Clinical Practice Research Datalink, CPRD); NPV: negative predictive value; PPV: positive predictive value; THIN: The Health Improvement Network

^a Results from several stratified analyses and from drug use in other pregnancy time windows are presented in the publication.

Authors' accepted version

TABLE 3. Methods Based on Timing of Prenatal Health Care

Publication	Study population	Data source	Estimation method	Comments	Publications which used related methods
Hardy, 2006 ¹⁹	Mapped pregnancies ⁴⁶ with ≤ 280 days between the earliest pregnancy marker (i.e., codes indicative of an ongoing pregnancy) and the pregnancy outcome (e.g., code for live birth), in women aged 15-44 years, with ≥ 7 months of prenatal records, linked to infants with ≥ 2 infant records in the first year of life. One pregnancy and one infant per woman were retained. These pregnancies were considered term or close to term.	1991-1999 Electronic medical records (GPRD), UK	The reference date was the date with a first indicator of pregnancy (visit, procedure, etc.). Two periods of exposure were assessed: 90 days before the reference date (intended to capture filled prescriptions for medications that were to be taken in the first trimester of pregnancy, or medications that would remain in the body early in pregnancy; it likely included the date of conception and the earliest weeks of pregnancy), and 70 days after the reference date (representing the first trimester).		Identification of early pregnancy markers: Piper, 1990 ⁴⁷ Manson, 2001 ⁵ Hardy, 2004 ⁴⁶ Devine, 2010 ⁴⁸
Kharbanda, 2012 ²⁰	Uncomplicated live births with birth certificate data in women enrolled in the Vaccine Safety Datalink.	2002-2010 Health care claims from several managed care organizations, USA	Pregnancies were hierarchically assigned a duration of pregnancy based on the presence of codes for 5 common prenatal tests and their assumed gestational timing.	20% of the pregnancies did not have such codes; it is not clear how these pregnancies were handled. The population on which the algorithm was developed and validated is not described in detail in this publication.	
Margulis, 2013 ⁸	Hospital live births in province's perinatal database with 365+280 days of outpatient coverage prior to the delivery date and gestational age at birth between 20 and 44 completed weeks.	1998-2007 Health care claims from single payer province-run health care system, Canada	Preterm deliveries were identified via a preterm status indicator. Within preterm and term deliveries as determined by the presence of the preterm status indicator, separately, linear regression was used to estimate the gestational age at birth based on the timing of prenatal screening tests.	Hospital records were needed to estimate coefficients; not needed if coefficients were applied to other populations	

Validation studies				
Publication	Study population	Data source	Methods	Results
Toh, 2008 ²⁶	3177 infants without malformations and their mothers	1998-2006 Slone Epidemiology Center Birth Defects Study, data from a single state, USA	This study replicated Hardy's ¹⁹ method. Exposure to antiinfectives in the first trimester (i.e., the first 70 days after the first prenatal visit) was estimated and compared to the exposure to the same drugs using Slone Epidemiology Center's information as the gold standard. Sensitivity and specificity of the estimated first-trimester exposure to antiinfectives were reported.	From 3177 women in the data source, 2447 were eligible following Hardy et al. eligibility criteria. The sensitivity (95% CI) for first-trimester exposure to antiinfectives was 59.1% (53.3%; 64.5%); the specificity was 98.1% (97.5%; 98.7%). When, instead of the original eligibility criteria, the 3177 women were included and stratified by term/preterm birth, the sensitivity for terms was 56.5% (51.1%; 61.7%); for preterms, it was 55.3% (38.3%; 71.4%). Specificity remained above 97% in all analyses.
Devine, 2008 ^{35 a}	417,946 live births	1987-2006 Electronic medical records (GPRD), UK	The date of last menstrual period was back calculated from 301,384 expected delivery dates. Pregnancies that had the expected delivery date estimated with high certainty ("sure" in the data field for expected delivery date certainty) or that had a date of last menstrual period close to the one in the additional clinical data were retained. Then the date of last menstrual period was estimated as 60, 50, 40 and 30 days prior to the first marker of pregnancy care for all deliveries.	52,073 physician-based dates of last menstrual period were identified and linked to 30,240 estimated dates. The absolute mean difference between the physician-based and the estimated date of last menstrual period was 16.0 days (95% CI 15.8; 16.1), 11.8 (11.7; 12.0), 12.7 (12.5; 12.9), and 20.0 (19.7; 20.2), for the 60, 50, 40, and 30 day estimations, respectively.
Kharbanda, 2012 ²⁰	See above	See above	The algorithm was validated with a sample of women from a single site. The validation methods are not described.	The mean (90% CI) for the difference in days between the estimated gestational age and the birth certificate's was 0.92 (0.5; 1.4). ">97% accurate in assigning vaccination exposure by trimester"; accuracy is not defined.

Margulis, 2013 ⁸	See above	See above	See above. Agreement of estimated duration of pregnancy and clinical gestational age at birth are reported stratified by gestational age at birth <37 (preterm deliveries) or ≥37 completed weeks (term deliveries) in hospital records.	Among preterm deliveries, 75.8% had an estimated duration of pregnancy within 2 weeks of the clinical gestational age at birth. Among term deliveries, the figure was 93.9%. Other methods based on visits for prenatal screening were evaluated and showed lower percentages of agreement.
-----------------------------	-----------	-----------	---	---

CI: confidence interval; GPRD: General Practice Research Database (currently Clinical Practice Research Datalink, CPRD)

^a Abstract

TABLE 4. Method Based on Birth Weight

Publication	Study population	Data source	Estimation method	Comments	Publications which used related methods
Piper, 1994 ²¹	White or black women who delivered a single live infant with a recorded birth weight of 500 to 6000 grams or a stillborn infant. Excluded women enrolled before the start of the index pregnancy. 1-2% of birth and fetal death certificates were excluded because of missing data.	1988-1989 Health care claims from Medicaid from a single state, USA	The date of last menstrual period from the birth or death certificate was considered the beginning of pregnancy. If only month and year were available, the date of last menstrual period was imputed as the 15 th day of the month. For deliveries with missing month or year, or live births with duration of pregnancy <140 or >294 days (15%), it was estimated as the median duration of pregnancy for infants of similar race and birth weight.	Requires availability of summary information on birth weight by gestational age at birth and race.	Piper, 1987 ⁴⁰ Ray, 1998 ⁴¹ Cooper, 2006 ⁴² Cooper, 2007 ³⁹

TABLE 5. Methods That Combine Elements From #2 and #3

Publication	Study population	Data source	Estimation method	Comments	Publications which used related methods
Hornbrook, 2007 ²²	Among 251,251 women 12-55 years old enrolled for at least 42 days in the study period, pregnancy episodes were identified from diagnosis (ICD-9), procedure, imaging, laboratory and pharmacy claims. Pregnancies were retained if the entire pregnancy took place within the study period and the woman was enrolled at the time of the pregnancy outcome (24,680 pregnancies in 21,001 women). Pregnancy outcomes included were: ectopic pregnancies, spontaneous abortions, therapeutic abortions, very preterm live birth, stillbirth, preterm live birth, term live birth.	1998-2001 Health care claims from a single managed care organization, USA	Pregnancy outcomes and their dates were identified. Then, a hierarchical algorithm was applied to estimate the date of the beginning or pregnancy. Preference for estimation of the gestational age at pregnancy end was given in this order: gestational age in the hospital discharge record, gestational age in alpha fetoprotein test, 40 weeks for pregnancies in the Preterm Birth Prevention Program database, gestational age from emergency department admission date, gestational age at outcome from national median figures.		Kharbanda 2012 ²⁰ Naleway, 2013 ²³
Snowball, 2007 ^{24 a}	Among 2,186,366 women considered valid members of GPRD aged 11 to 49, pregnancies were identified based on codes related to antenatal, neonatal and postnatal care, pregnancy, birth and termination. Pregnancy outcomes (deliveries and terminations) and their dates were identified (494,449 pregnancies of which 359,712 had delivery for outcome).	1992-2006 Electronic medical records (GPRD), UK	A hierarchical algorithm assigned pregnancy duration based on (in this order) a record of the estimated due date, a record of the first day of the last menstrual period, a code specifying gestational age (e.g., "baby premature 36 weeks"), or a code specifying pre or post-term delivery without reference to a specific gestational age (defaults values were 36 and 41 weeks, respectively). Pregnancies without any of the above were assigned a default duration of 40 weeks for deliveries, or 10 weeks for terminations.		Charlton, 2011 ²⁵ Sammon, 2012 ⁴⁵ Ban, 2012 ⁴³ Charlton, 2013 ⁴⁴
Validation studies					

Publication	Study population	Data source	Methods	Results
Hornbrook, 2007 ²²	See above	See above. The validation sample included 511 women; pregnancies were randomly selected for validation within pregnancy outcome strata (678 pregnancies)	Information abstracted from medical charts: presence of a pregnancy, pregnancy outcome, and pregnancy outcome and beginning dates. Agreement for outcome dates was considered present when the algorithm and medical chart outcome dates were within 30 days of each other. Among those with agreement, gestational age agreement within 2 weeks and 15-28 days was estimated.	94% of pregnancies (640/678) agreed on the outcome date. 81% of the 640 pregnancies agreed on gestational age at birth within 2 weeks, and 13% agreed within 15-28 days. Agreement by pregnancy outcome and other stratified results are presented in the original publication.
Naleway, 2013 ²³	For the validation study, 420 pregnancies enrolled in the Vaccine Safety Datalink were sampled (15 pregnancies ending in: live birth, spontaneous abortion, elective or therapeutic abortion, and other pregnancy outcomes). Eligible pregnancies were those ending in 2002-2006 in women 12-55 years old continuously enrolled from 3 months prior to pregnancy through 1 month after the end of pregnancy. A single pregnancy per woman was retained.	2002-2006 Electronic medical records, health care claims and other sources from several health care plans, USA	Hornbrook's algorithm was adapted for use with the information available in the data source (laboratory tests, pharmacy dispensing and imaging procedures were not available). Agreement for pregnancy end date and for gestational age at the end of pregnancy was considered present when the algorithm- and the medical-chart-derived numbers were within 30 days of each other.	In 67% of all pregnancies, there was an exact match between the algorithm and the medical chart pregnancy end dates; in 84% of pregnancies, end dates from the two sources were within 14 days of each other. Among pregnancies ending in live births, agreement on the pregnancy end date was present in 96% of the pregnancies; agreement on gestational age at birth was present in 98% of the subset with recorded gestational age at birth in the chart. Among pregnancies ending in spontaneous abortions, the figures were 95% and 83%, respectively. Among pregnancies ending in therapeutic abortion, the figures were 70% and 85%, respectively. Among pregnancies ending in other outcomes, figures were 85% and 82%, respectively.

^a The algorithm described here was presented in detail in the associated poster at the 2007 International Conference on Pharmacoepidemiology and Therapeutic Risk Management.